Retail Sales & Customer Behaviour Analysis

1. Project Overview

In today's competitive retail environment, data-driven decision-making plays a critical role in improving sales performance and understanding customer behavior. DataVista Pvt. Ltd. was engaged by a mid-sized retail organization to analyze customer purchasing patterns, product performance, and overall revenue trends. The objective of this project is to leverage historical transaction data to identify actionable insights that can support strategic business decisions, optimize inventory planning, and drive sustainable revenue growth.

2. Dataset Description

The analysis was performed using four structured datasets provided by the organization:

- **customers.csv**: Contains customer demographic information such as CustomerID, CustomerName, City, Age, Gender, and SignupDate.
- **products.csv**: Includes product-related details such as ProductID, ProductName, Category, and UnitPrice.
- **orders.csv**: Stores order-level information, including OrderID, CustomerID, OrderDate, and PaymentMode.
- **order_items.csv**: Captures item-level transaction details such as OrderItemID, OrderID, ProductID, Quantity, and DiscountPercent.

The datasets are related through primary and foreign keys, enabling a comprehensive view of customer transactions. Product names were anonymized in the dataset to maintain confidentiality and comply with data privacy requirements.

3. Data Cleaning & Preparation

The initial phase of the project focused on understanding and preparing the datasets for analysis. All datasets were examined for missing values, duplicate records, and data inconsistencies. No missing or duplicate records were identified, ensuring high data quality.

Date columns such as SignupDate and OrderDate were converted into a datetime format to support time-based analysis. The datasets were then merged using appropriate primary and foreign keys to create a unified analytical dataset.

A calculated field, **Revenue**, was created using the following formula:

Revenue = Quantity × UnitPrice × (1 − DiscountPercent / 100)

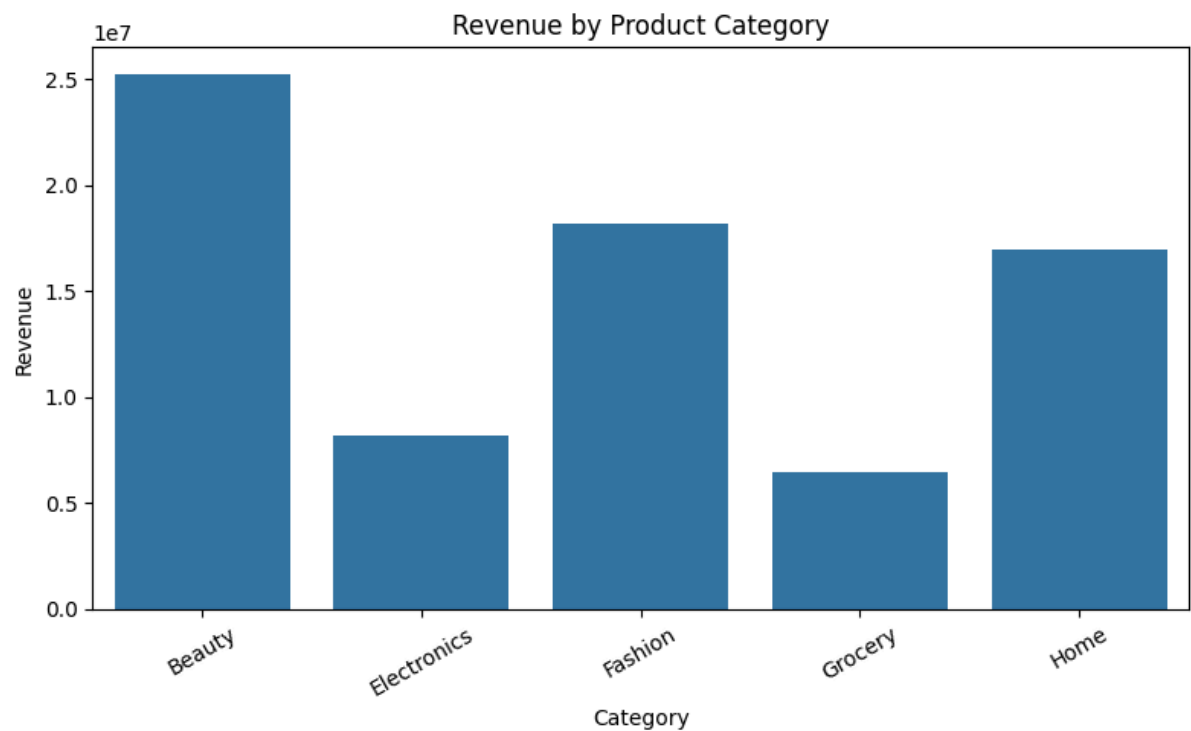This revenue metric served as the foundation for all subsequent analysis and visualizations.
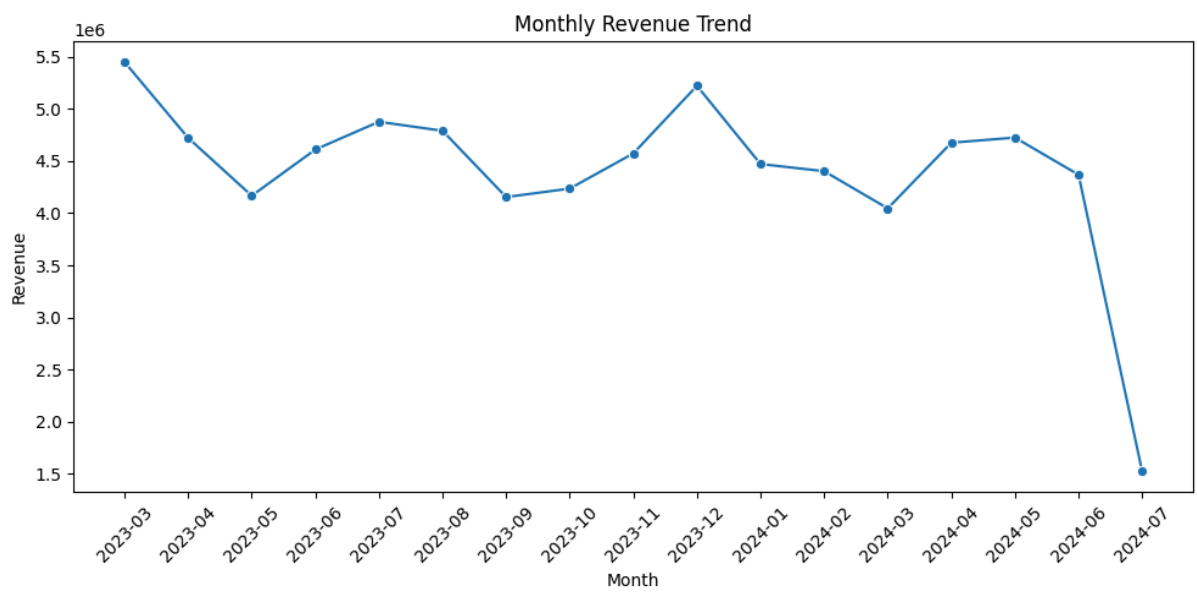
---

4. Exploratory Data Analysis (EDA)

Exploratory data analysis was conducted to identify trends, patterns, and relationships within the data. Key performance indicators such as total revenue, monthly sales trends, top-selling products, and payment mode preferences were analyzed.
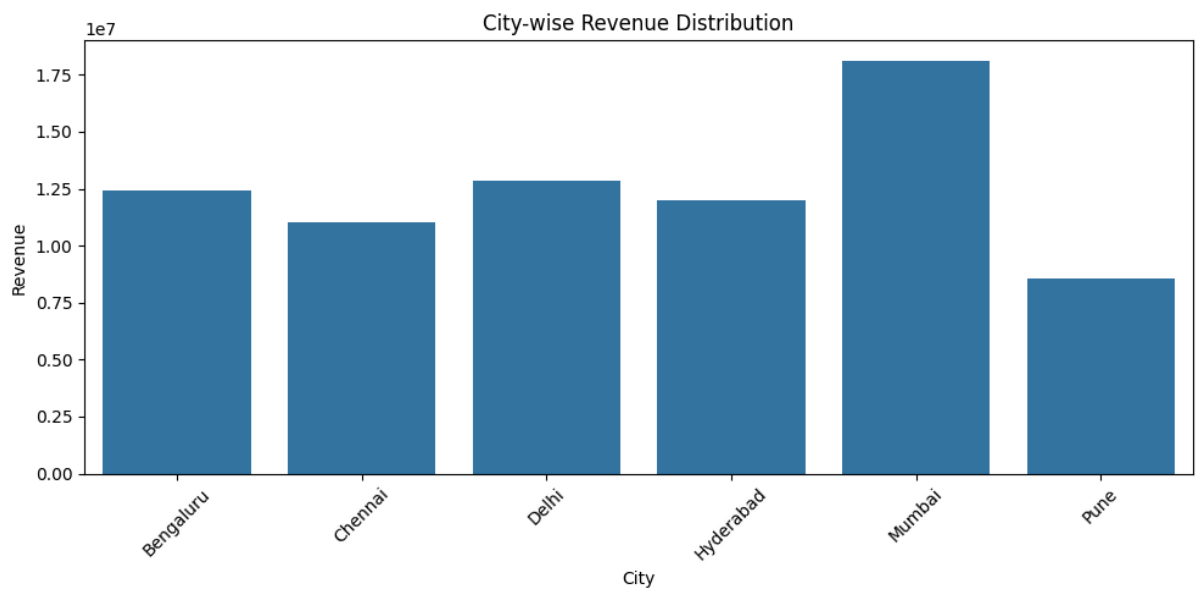
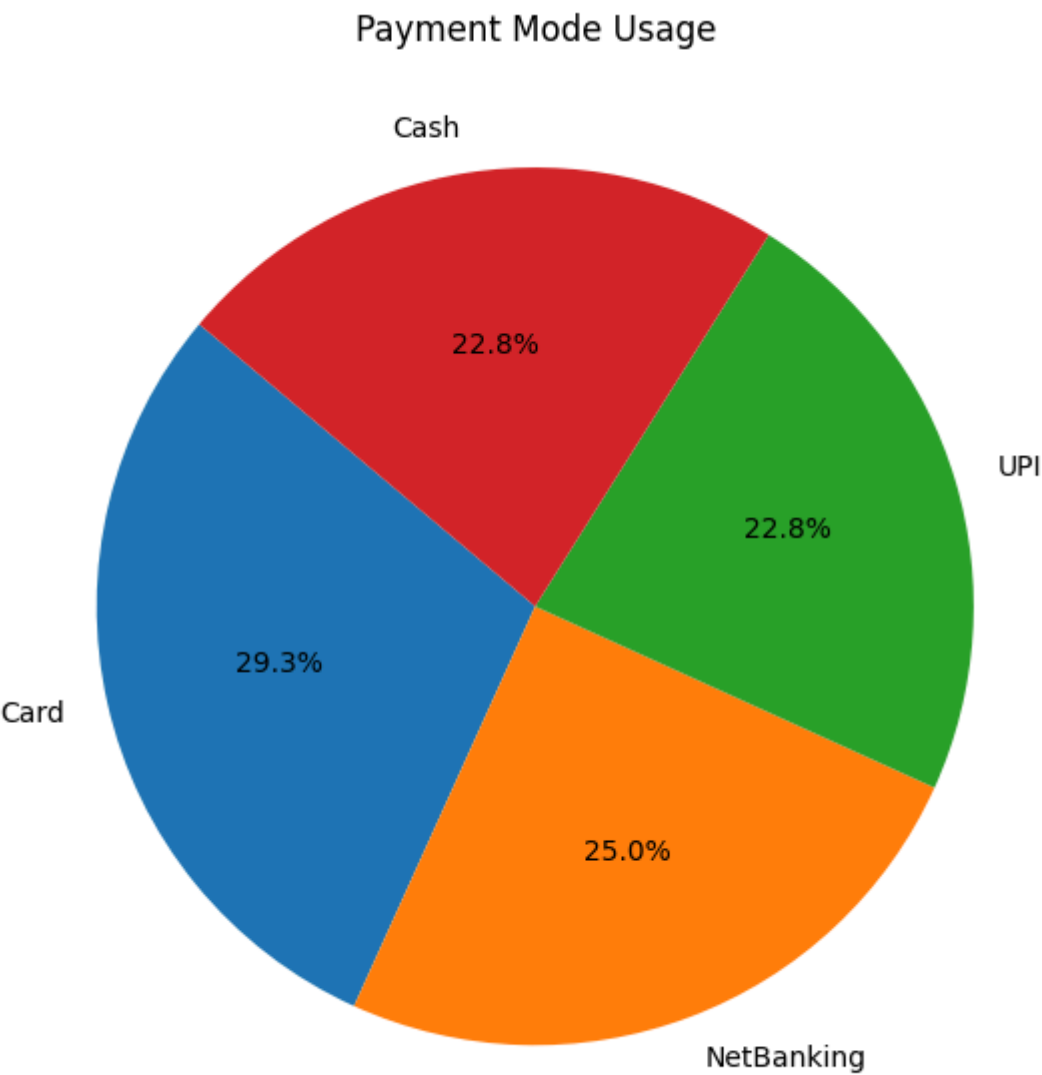## Visualizations Included:

- **Revenue by Category (Bar Chart)**
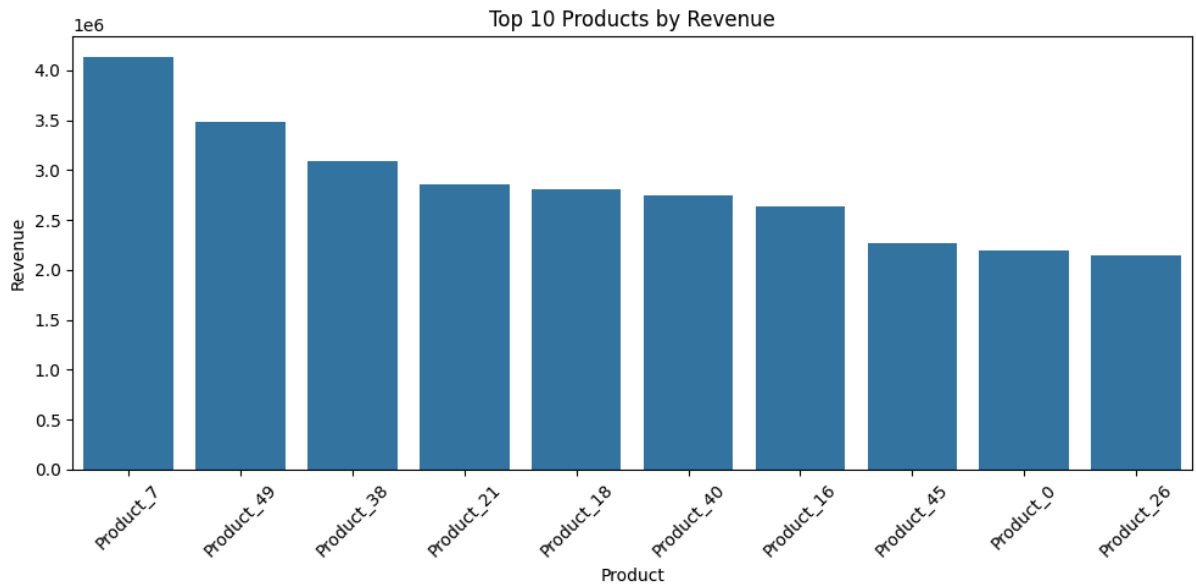
● **Monthly Revenue Trend (Line Chart)**

● **City-wise Revenue Distribution (Bar/Pie Chart)**

- **Payment Mode Usage (Pie Chart)**

Payment Mode Usage

- **Top 10 Products by Revenue (Bar Chart)**

Top 10 Products by Revenue

These visualizations provided a clear overview of sales performance across categories, time periods, and geographic regions.

---

5. Key Insights & Findings

Based on the exploratory and customer-level analysis, the following key insights were identified:

- A limited number of cities contribute a significant portion of total revenue, indicating concentrated regional demand.
- Certain product categories, such as Electronics and Fashion, outperform others in terms of revenue generation.
- Returning customers place multiple orders and contribute higher overall revenue compared to one-time buyers.
- Specific age groups dominate the customer base, highlighting the primary target audience.
- Digital payment methods are more widely used than cash-based transactions.
- Several products consistently generate low revenue, increasing inventory holding Risk.

---

6. Business Insights & Recommendations

6.1 Key Business Insights

The analysis highlights strong opportunities for optimizing marketing, pricing, and inventory strategies. Revenue concentration across select cities and categories suggests that focused efforts can yield significant business impact. Discount strategies influence sales volume, but excessive discounts may erode profit margins.

6.2 Actionable Recommendations

- **City-Level Strategy**: Prioritize marketing campaigns and operational improvements in top-performing cities to maximize return on investment.
- **Category Optimization**: Increase inventory availability and promotional efforts for high-performing product categories while reviewing underperforming categories.
- **Discount Strategy**: Implement targeted discounting for slow-moving or seasonal products rather than uniform high discounts across all products.
- **Inventory Management**: Review low-selling products for potential discontinuation or replacement to reduce inventory risk.
- **Customer Retention**: Introduce loyalty programs and personalized offers to encourage repeat purchases and improve customer lifetime value.
- **Payment Optimization**: Promote preferred digital payment modes through incentives to improve transaction efficiency and customer experience.

7. Conclusion

This project demonstrates how data analytics can provide valuable insights into customer behavior, product performance, and revenue trends in the retail sector. The findings offer practical recommendations that can support informed decision-making and long-term business growth. Future enhancements may include predictive sales forecasting, customer lifetime value analysis, and advanced segmentation Techniques, to further strengthen business strategy.